JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

# PROJECT DELIVERABLE REPORT

| | |
|---|---|
| DELIVERABLE NUMBER AND TITLE | D2.4.1 |
| TITLE | CASE STUDIES: PREPARING JOINT DATA ANALYSIS |
| AUTHOR(S) | DR. G. PEROZZI, DR. R. CANALI, DR. L.O. DRAGSTED, S. PALOMBI, M. TULLIN, DR. M. PINART, DR. T PISCHON, DR. J. BOUWMAN, DR.K. NIMPTSCH |
| WORK PACKAGE | WP 2 |
| TASK | TASK 2.4 |
| WP LEADER | DR. LARS OVE DRAGSTED |
| BENEFICIARIES CONTRIBUTING TO THE DELIVERABLE | CREA-NUT, UCPH, MDC, TNO |
| STATUS – VERSION | DRAFT - VERSION 1.0 |
| DELIVERY DATE (MONTH) | M30 |
| SUBMISSION DATE | SEPT 8, 2017 |
| DISSEMINATION LEVEL – SECURITY* | PU |
| DELIVERABLE TYPE** | R |

* Security: PU – *Public*; PP – *Restricted to other programme participants (including JPI Services)*: RE – *Restricted to a group specified by the consortium (including JPI Services)*; CO – *Confidential, only for members of the consortium (including JPI Services)*

** Type: R – *Report*; P – *Prototype*; D – *Demonstrator*; - O - *Other*

# Table of contents

## CASE STUDIES DESIGN

To perform case studies aimed at conducting joint analysis of multiple studies through the DASH-IN infrastructure, a subset of the datasets available for upload in DASH-IN was selected on the basis of:

a) a research question that could be answered by merging and querying the appropriate datasets
b) a sufficient number of datasets containing the appropriate phenotypic measurements and available for sharing

### IDENTIFICATION OF A RESEARCH QUESTION

The research question on the effect of dietary components in health maintenance was based on the available scientific knowledge.

Cardiovascular diseases (CVD) are among the main causes of death in the world, and atherosclerosis plays a key role in their development, initiating as a vascular endothelium dysfunction. Atherosclerosis is triggered by several risk factors, including dyslipidemias (Mallika et al, 2007), which is characterized by increased plasma cholesterol, triglycerides (TGs), and/or high LDL/low HDL levels.

Diet was recognized to play a crucial role in the control of chronic diseases, including CVD. Primary (genetic) and secondary (lifestyle and other) causes contribute to dyslipidemias to varying degrees. The most important secondary cause in developed countries is sedentary lifestyle with excessive dietary intake of saturated fat, cholesterol, and trans-fatty acids. Food components have the potential to play a key role in the treatment and primary prevention of dyslipidemias and thereby in the control of CVD.

Several studies have shown the effectiveness of certain foods in reducing disease risk, and specific foods and food components can be employed as support strategies towards reducing the risk of dyslipidemia. They are consumed as part of the normal diet, exerting beneficial physiological effects and/or reducing the risk of chronic diseases, in addition to their basic nutritional functions. LDL cholesterol is a major modifiable risk factor for CVD with dietary and lifestyle changes and a combination of both was shown as the most effective preventive strategy, highlighting the tight relationship between these two factors (Joosten et al, 2010). However, this joint effect was rarely investigated experimentally and the significance of the available results was often limited by small sample size. Evidence from randomized controlled trials indicated that macronutrient composition influences lipoprotein profiles: replacement of carbohydrates with unsaturated fat was shown to decrease LDL and increase HDL cholesterol (Miller et al, 2006). On the other hand, replacement of carbohydrates with saturated fat did not affect lipoprotein profiles in a meta-analysis of randomized controlled trials (Mensink et al, 2003).

Growing interest in the cholesterol-reducing effects of soluble dietary fibres was also reported over the past 50 years, and a general consensus has now been reached on the capacity of fiber to lower circulating cholesterol concentrations. However, the efficacy of cholesterol lowering is most likely a function of fiber viscosity, and therefore fiber type. Soluble fiber is usually highly viscous and fermented by the intestinal microflora. It is mainly composed of pectin (found in fruits), gums (oats, barley, and legumes such as soybeans), and mucilage (Zhu et al, 2015; Khalesi et al, 2016; Ho et al, 2017). Three major biological mechanisms have been proposed to explain the cholesterol-reducing effects of soluble fiber: prevention of the

re-absorption of bile salts from the small intestine; reduced glycemic response leading to lower insulin stimulation of hepatic cholesterol synthesis; and physiological effects of fermentation products, mainly propionate.

On the basis of the above knowledge, we thought it would be important to design case studies addressing the role of diet and physical activity on lipid metabolism and CVD biomarkers. This overarching question could be separately applied to both observational and intervention studies, and drive the selection of the corresponding datasets to carry out two parallel case studies. In particular:

- nutritional epidemiological observation studies could greatly contribute to further investigate the association of macronutrient composition and physical activity with lipoprotein profile by joint data analysis.
- Intervention studies with potential cholesterol-lowering food components could be jointly analyzed for possible mechanistic insights on specific dietary components affecting lipid metabolism

## SELECTION OF CASE STUDIES DATASETS

The list of datasets available for integration in ENPADASI, as well as the main features of each study, were described in deliverable D2.1.1, which represented the starting point for developing case studies. Overall, the ENPADASI project partners committed to contribute existing datasets from 24 observational and 79 intervention studies, although some of these studies were completed and shared towards the end of the project and therefore were not initially included in D2.1.1.

As mentioned in D2.1.1, and illustrated in more detail by the results of a survey conducted among ENPADASI partners as part of Task 2.4 activities (Annex1), the major obstacles to sharing data among ENPADASI project partners were based on personal and/or technical issues. Therefore, selection of datasets for the case studies could include only those whose experimental results that had been uploaded in DASH-IN, while excluding all studies in which only metadata were available for sharing.

The main difference between observational and intervention studies lies in study designs and phenotypic measurements, which in turn lead to different technical issues in handling datasets for joint data analysis. This aspect was discussed in depth during dedicated skype meetings and led us to identify the optimal solution in designing two separate case studies, one including only nutritional epidemiological datasets and the other including datasets from nutritional intervention studies.

## CASE STUDY 1 - OBSERVATIONAL STUDIES

The ENPADASI partners contributing datasets from nutritional epidemiological studies, potentially available for sharing, were initially requested to fill-in a template containing metadata on study design, exposure measurements (dietary intake, alcohol and tobacco consumption, physical activity, sedentary behaviour, anthropometry, sociodemographic and health status), main health-related outcomes, and laboratory measurements (clinical and omics biomarkers). To increase participation, their feedback was requested not only on the studies initially listed as partner contribution to the project (full list in D.2.1.1), but was also extended to any other study to be completed during the ENPADASI project lifetime and complying with the

inclusion criteria (see above described research question). This work led to identifying 26 observational studies which could be used for the observational case study: 12 cohort, 12 cross-sectional and 2 case-control design, accompanied by the appropriate phenotypic measurements. In particular, 20 of the datasets included measurements of clinical biomarkers (lipoproteins, glucose, insulin, markers of inflammation), while 12 datasets also included –omics data. The selection process as well as the full details of each of the selected dataset are reported in a Manuscript submitted for publication (Pinart M *et al*, "Identification of observational studies with nutritional epidemiological data within ENPADASI" submitted to the *Public Health Nutrition* journal).

## CASE STUDY 2 - INTERVENTION STUDIES

The 79 nutritional intervention studies contributed by ENPADASI project partners and potentially available for sharing, were initially catalogued by requesting to fill-in the template described in D2.1.1, listing metadata on the study design (intervention type, subjects, inclusion/exclusion criteria), phenotypic quantifications (clinical chemistry, -omic measurements) and availability to share datasets. Due to the intrinsic complexity of intervention studies, many of them were conducted using animal models (rodents), *in vitro* cell cultures or *ex-vivo* treatments of human primary cells, while a smaller fraction of interventions had been conducted on human subjects. Joint analysis of data from such different study designs cannot be performed without loosing biological significance. We therefore chose to limit the selection of case study datasets to studies involving dietary interventions on human subjects with phenotypic outcomes measuring clinical and –omic parameters of lipid metabolism. Of the 19 selected studies, 10 were carried out with a parallel design, 8 with a Crossover scheme and 1 employed both designs. The dietary intervention included whole diets (9 studies), single foods or supplements (6 studies) or a short term dietary challenge (OGGT or OLTT, 3 studies). Duration of the intervention spanned from 4 weeks to 24 months, and food intake was assessed in all cases, using a Food Frequency Questionnaire, or 24h recall (11 studies). 7 studies report having provided meals (or supplements) to the subjects for the entire duration of the study. Clinical parameters related to lipid metabolism were measured in all cases, and -omic phenotyping was mainly based on plasma/urine metabolomics.

# PREPARATION FOR JOINT DATA ANALYSIS

## CASE STUDY 1 - OBSERVATIONAL STUDIES

Joint data analysis of the combined datasets from observational studies was carried out using the DataSHIELD initiative (Gaye et al, 2015), which enables to perform data analysis without providing any direct access to individual-level data. This method allowed to circumvent the legal and ethical constraints preventing datasharing of raw data.

The aim of case study 1 was to investigate the joint association of macronutrient composition and physical activity with lipoprotein profile. We circulated a research proposal based on the metadata provided by the identified observational studies (n= 26) willing to share their (meta)data in ENPADASI. For our purposes, 12 observational studies (7 cross-sectional and 5 cohort studies) had the required data on lipids, macronutrients, and phyical activity (see Research proposal). All of them agreed to participate in the case study. One cross-sectional study was added though it was not initially part of ENPADASI but the collaborators were interested to join in. A variable catalogue was created and circulated to the 13 studies to check for data availability. Twelve studies had the specific variables or had other variables avaialble to construct the requested ones. One longitudinal study did not have some of the required variables and was left out. The remaining 12 studies listed in Table 1 were requested to build a harmonized database to be uploaded in their OPAL server, which is the management interface of DataSHIELD, where the data is kept. They were asked to circulate their data dictionary to the ENPADASI partner leading the case study (Max Delbrück Center (MDC) for Molecular Medicine. So far, 7 studies have their harmonized dataset uploaded into OPAL and sent their data dictionary. The next step is to have the access to their datasets via the federated approach. Briefly, a central computer will send the syntax in R to the OPAL servers and a summary statistics will be retrieved.

The central computer is located in MDC, where they are currently dealing with security issues in order to be able to start analysing the data from 5 studies whose data are already available (Table 1).

**Table 1. List of the participating studies**

| Study ID | Study design | Studies ready for analysis* |
|---|---|---|
| EPIC substudy | Cohort | Yes |
| INGI-FVG | Cohort | No |
| GINI-LISA | Cohort | No |
| DONALD study | Cohort | Yes |
| NESCaV | Cross-sectional | Yes |
| NVSII | Cross-sectional | Yes |
| BVSII | Cross-sectional | Yes |
| Pizarra | Cross-sectional | No |
| Cardiovascular and metabolic syndrome risk assessment of Bolivian adolescents | Cross-sectional | No |
| Metbanc | Cross-sectional | No |
| NANS | Cross-sectional | No |
| ActivE study | Cross-sectional | No |

*Studies whose dataset is uploaded into OPAL AND the data dictionary is provided AND have succesfully installed DataSHIELD (have provided their login and passwords)

In previous cross-sectional analyses, the joint effect of macronutrient composition and physical activity was rarely investigated and the ability to detect joint and interaction effects was sometimes limited by low sample sizes. Combining data from twelve European observational studies will enhance statistical power to detect interactions and allow performance of subgroup analyses by sex, age groups and geographical region.

## CASE STUDY 2 - INTERVENTION STUDIES

The selected datasets were uploaded into the Phenotype Database (studies.dbnp.org) for joint data analysis. Six of these allowed full datasharing, while the remaining 5 allowed sharing of metadata alone (Table 2)

**Table 2. List of the intervention studies datasets shared in DASH-IN**

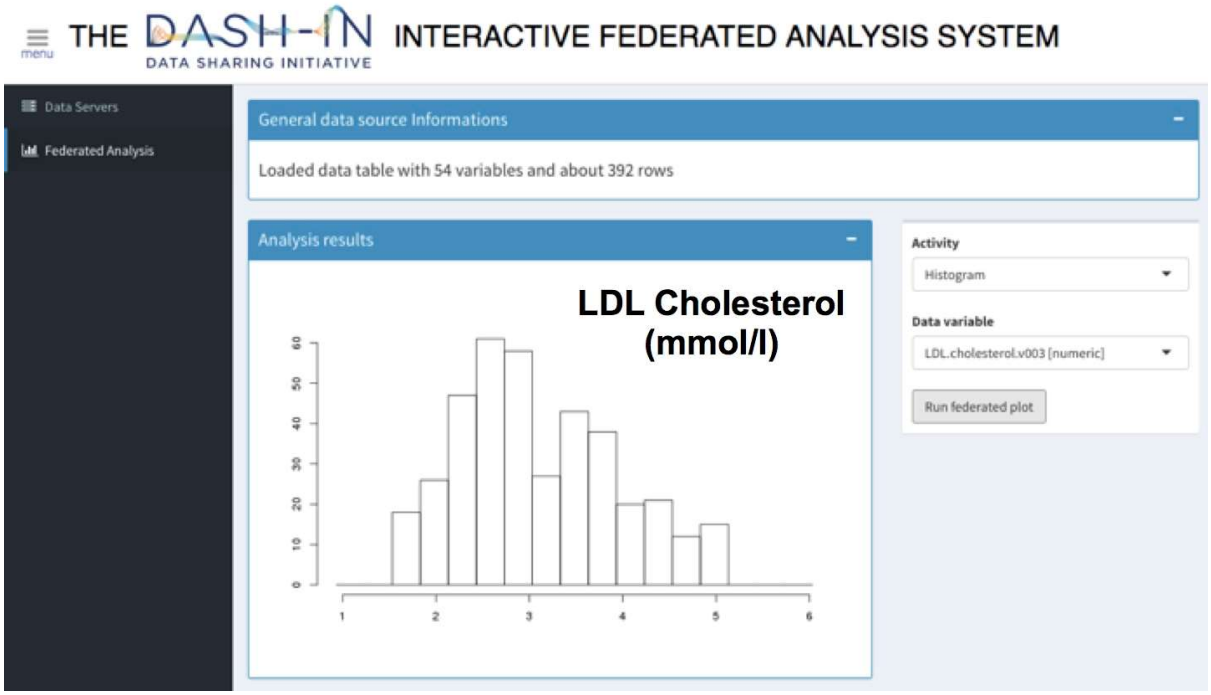
| Study ID and title | Study design/ subjects | Exposure | Endpoints |
|---|---|---|---|
| ADMIT – Assessment of Dietary Modulation of Inflammatory Tone | Cross-over 15 Subjects | Glucose (± lipids)/water | Physiology, Clinical chemistry, Proteomics |
| Diclofenac | Parallel 19 Subjects | Diclofenac/placebo followed by OGTT challenge | Physiology, Clinical chemistry, Transcriptomics, Proteomics, Metabolomics |
| Foodmix – Effect of Nutritional Interventions on Inflammatory Status in Healthy Overweight Man | Partial Cross-over 42 Subjects | Anti-inflammatory Food components/probiotics/placebo | Physiology, Clinical chemistry, Transcriptomics, Proteomics, Metabolomics |
| NuGO_PPSH – Human Proof of Principle Study: an intervention study | Cross-over 10 Subjects | 12h or 36h fasting/repeatability of 12h fasting | Clinical chemistry, Transcriptomics, Proteomics, plasma and urine Metabolomics |
| LIPGENE: Effects of dietary fat modification on insulin sensitivity and on other risk factors of the metabolic syndrome | Parallel 416 Subjects | High fat SFA rich diet/ High fat MUFA rich diet/Low fat High complex Carbohydrate isoenergetic diet/ Low fat High complex Carbohydrate isoenergetic diet with VLC n-3 PUFA | Clinical chemistry |
| HuMet_mrt_2012 – The dynamic range of the human metabolome revealed by challenges | Sequential challenges 15 Subjects | Challenge tests: 36h fasting, exercise, OGTT, OLTT, cold stress, defined liquid meals | Clinical chemistry, NMR metabolomics |
| 9218 Fat challenge tests | Parallel and sequential 20 Subjects | OLTT challenge and 2 wks high fat intervention | Physiology, Clinical chemistry, Transcriptomics, Metabolomics, Microbiota, Questionnaire |
| A308 Cheese – Human dietary intervention with cheese *vs* butter | Cross-over 23 Subjects | 6 wks equicaloric diets with cheese or butter | Clinical chemistry, Antropometrics, Metabolomics |
| M208 Berries – Human dietary challenge with strawberry, sea buckthorn or water, together with | Cross-over 18 Subjects | strawberry, sea buckthorn or water, together with a sugar | Clinical chemistry, Antropometrics, Metabolomics, Microbiota |

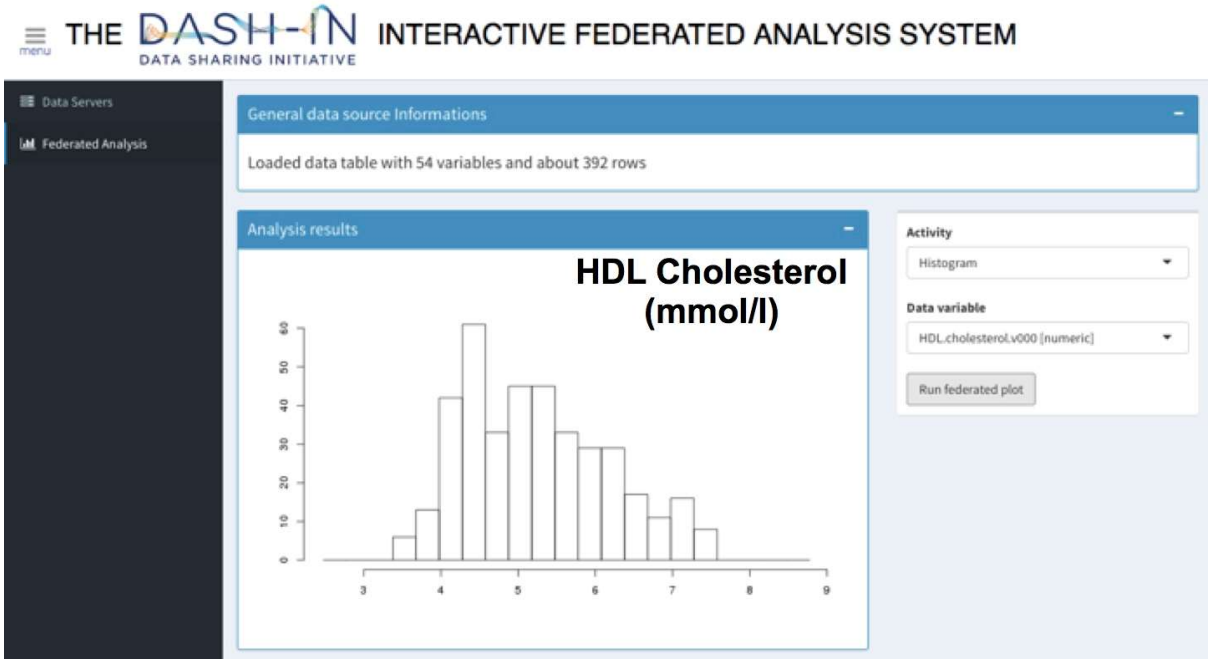
| | | | |
|---|---|---|---|
| a sugar load | | load | |
| M187 Farm-trout – The effects of trouts fed with a vegetable based feed on cardiovascular risk markers and plasma proteome | Parallel<br>68 Subjects | 3 diets fed to farmed trouts: vegetable based feed/marine feed/chicken 8wks feeding of trouts to human subjects | Clinical chemistry, Antropometrics, plasma and urine Metabolomics, Microbiota, Questionnaire |
| M143 six-a-day – Effects of fruit and vegetable on surrogate markers of oxidative damage to DNA, protein and lipids, enzymic defence and lipid metabolism determined in blood and urineParallel | Cross-over<br>43 Subjects | 3 diets: Fruit-Vegetable/Vitamin pill/ placebo | Clinical chemistry |

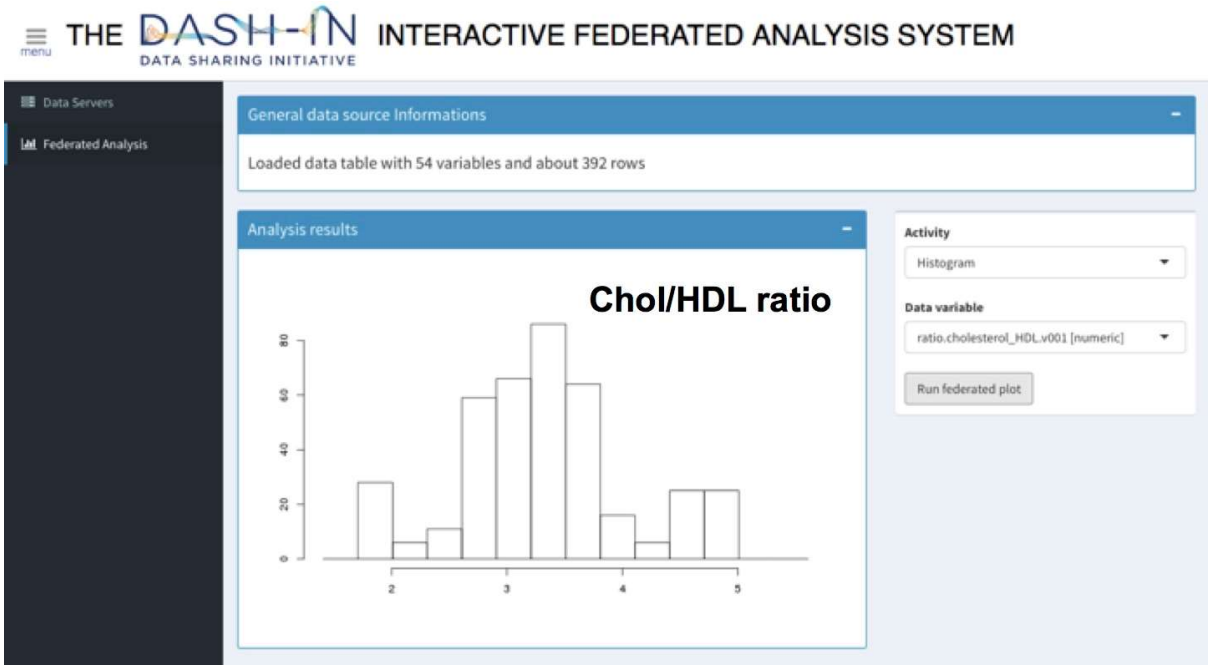
## PROOF OF PRINCIPLE: FEDERATED DATA ANALYSIS

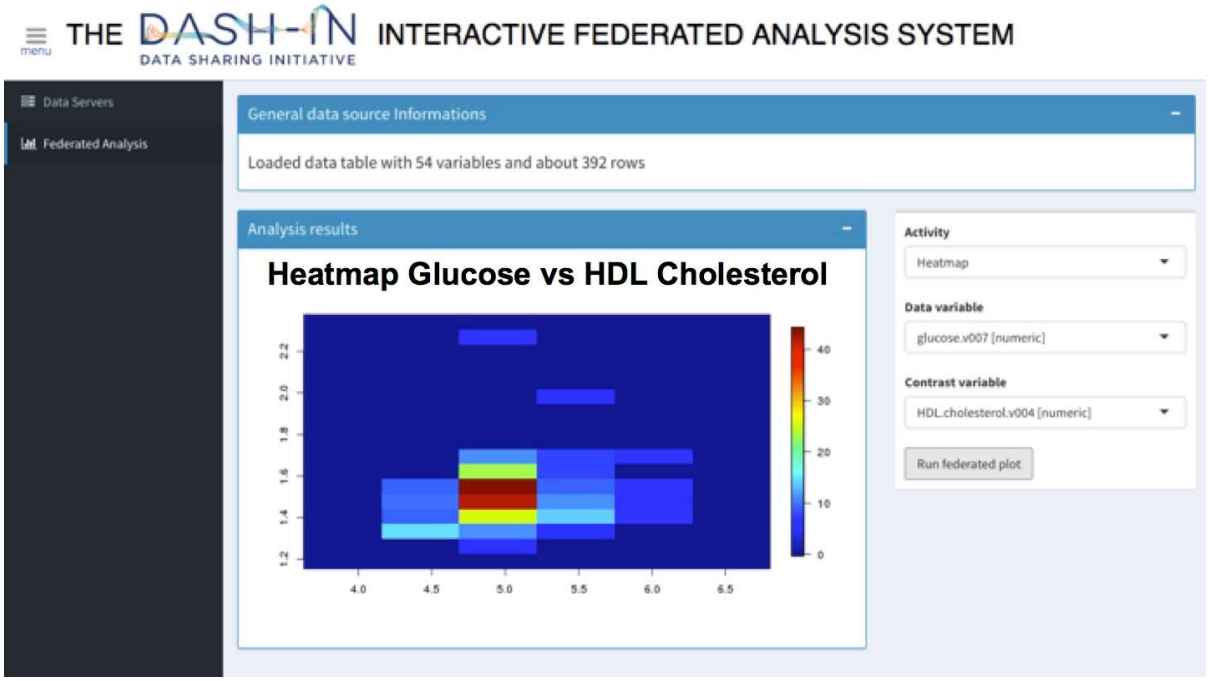
Of the 12 studies sharing full data, clinical biomarkers of lipid metabolism were measured in all studies, while 8 of them also included metabolomics datasets, 4 of them Transcriptomics data, 4 Proteomics, 2 Microbiota profiles. The lipid metabolism and glucose metabolism data from these studies have been employed to test the DASH-IN tool for joint data analysis developed within WP3 of the ENPADASI project.

The resulting plots for the ADMIT study are shown below:




THE DASH-IN DATA SHARING INITIATIVE INTERACTIVE FEDERATED ANALYSIS SYSTEM
menu
Data Servers
Federated Analysis
General data source Informations
Loaded data table with 54 variables and about 392 rows
Analysis results
HDL Cholesterol (mmol/l)
Activity
Histogram
Data variable
HDL.cholesterol.v000 [numeric]
Run federated plot


THE DASH-IN DATA SHARING INITIATIVE INTERACTIVE FEDERATED ANALYSIS SYSTEM
menu
Data Servers
Federated Analysis
General data source Informations
Loaded data table with 54 variables and about 392 rows
Analysis results
Chol/HDL ratio
Activity
Histogram
Data variable
ratio.cholesterol_HDL.v001 [numeric]
Run federated plot




## CONCLUSIONS

The work carried out within Task 2.4 of the project allowed to identify 12 observational studies and 12 intervention studies which can be used in two separate case studies for joint data analysis. The major limitation emerging from this work was the difficulty of study owners to allow datasharing of the full datasets, but this trend will likely change after the ENPADASI project, as other WPs have delivered solutions to the ethical and legal issues that represented major obstacles. Despite this issue, joint data analysis of 12 intervention study datasets was carried out using different parameters of lipid and glucose metabolism from the selected studies in Table 2, demonstrating successful development of the powerful ENPADASI tool for federated data analysis.

## REFERENCES

Gaye A, Marcon Y, Isaeva J et al. (2014) DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol 43, 1929-1944

Ho HVT, Jovanovski E, Zurbau A, Blanco Mejia S, Sievenpiper JL, Au-Yeung F, Jenkins AL, Duvnjak L, Leiter L, Vuksan V. A systematic review and meta-analysis of randomized controlled trials of the effect of konjac glucomannan, a viscous soluble fiber, on LDL cholesterol and the new lipid targets non-HDL cholesterol and apolipoprotein B. Am J Clin Nutr. 2017 May;105(5):1239-1247

Joosten MM, Grobbee DE, van der AD, Verschuren WM, Hendriks HF, Beulens JW. Combined effect of alcohol consumption and lifestyle behaviors on risk of type 2 diabetes. The American journal of clinical nutrition 2010;91:1777-83

Khalesi S, Paukste E, Nikbakht E, Khosravi-Boroujeni H. Sesame fractions and lipid profiles: a systematic review and meta-analysis of controlled trials. Br J Nutr. 2016 Mar 14;115(5):764-73

Mallika V, Goswami B, Rajappa M. Atherosclerosis pathophysiology and the role of novel risk factors: a clinicobiochemical perspective. Angiology 2007;5 8:513–22).

Mensink RP, Zock PL, Kester AD, Katan MB. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. The American journal of clinical nutrition 2003;77:1146-55.

Miller ER, Erlinger TP, Appel LJ. The effects of macronutrients on blood pressure and lipids: an overview of the DASH and OmniHeart trials. Current atherosclerosis reports 2006;8:460-5.

Zhu X, Sun X, Wang M, Zhang C, Cao Y, Mo G, Liang J, Zhu S. Quantitative assessment of the effects of beta-glucan consumption on serum lipid profile and glucose level in hypercholesterolemic subjects. Nutr Metab Cardiovasc Dis. 2015 Aug;25(8):714-23